Virology **291,** 226–234 (2001)
doi:10.1006/viro.2001.1190, available online at http://www.idealibrary.com on IDEAL®

Sequences and Replication of Genomes of the Archaeal Rudiviruses SIRV1 and SIRV2: Relationships to the Archaeal Lipothrixvirus SIFV and Some Eukaryal Viruses[1]

Xu Peng,* Helmut Blum,† Qunxin She,* Sylvia Mallok,† Kim Brügger,* Roger A. Garrett,* Wolfram Zillig,‡ and David Prangishvili§,[2]

*Microbial Genome Centre, Institute of Molecular Biology, Copenhagen University, Sølvgade 83H, DK-1307 Copenhagen K, Denmark; †Genzentrum der Ludwig-Maximilians-Universität München, Feodor-Lynen-Strasse 25, D-81377 Munich, Germany; ‡Max-Planck-Institute für Biochemie, Am Klopferspitz 18, D-82152 Martinsried, Germany; and §Lehrstuhl für Mikrobiologie-Archaeenzentrum, Universität Regensburg, Universitätsstrasse 31, D-93053 Regensburg, Germany

Received May 8, 2001; returned to author for revision August 15, 2001; accepted September 10, 2001

The double-stranded DNA genomes of the viruses SIRV1 and SIRV2, which infect the extremely thermophilic archaeon *Sulfolobus* and belong to the family *Rudiviridae,* were sequenced. They are linear, covalently closed at the ends, and 32,312 and 35,502 bp long, respectively, with an A+T content of 75%. The genomes of SIRV1 and SIRV2 carry inverted terminal repeats of 2029 and 1628 bp, respectively, which contain multiple direct repeats. SIRV1 and SIRV2 genomes contain 45 and 54 ORFs, respectively, of which 44 are homologous to one another. Their predicted functions include a DNA polymerase, a Holliday junction resolvase, and a dUTPase. The genomes consist of blocks with well-conserved sequences separated by nonconserved sequences. Recombination, gene duplication, horizontal gene transfer, and substitution of viral genes by homologous host genes have contributed to their evolution. The finding of head-to-head and tail-to-tail linked replicative intermediates suggests that the linear genomes replicate by the same mechanism as the similarly organized linear genomes of the eukaryal poxviruses, African swine fever virus and *Chlorella* viruses. SIRV1 and SIRV2 both contain motifs that resemble the binding sites for Holliday junction resolvases of eukaryal viruses and may use common mechanisms for resolution of replicative intermediates. The results suggest a common origin of the replication machineries of the archaeal rudiviruses and the above-mentioned eukaryal viruses. About 1/3 of the ORFs of each rudivirus have homologs in the *Sulfolobus* virus SIFV of the family *Lipothrixviridae,* indicating that the two viral families form a superfamily. The finding of inverted repeats of at least 0.8 kb at the termini of the linear genome of SIFV supports this inference.

Key Words: Archaea; virus; genome; replication; recombination; virus evolution.

INTRODUCTION

Since the Archaea were discovered as the third domain of life (Woese and Fox, 1977), considerable insight has been gained into their biology (reviewed in Pfeifer *et al.,* 1994) and this process has been enhanced, recently, by the availability of several archaeal genome sequences. Moreover, several extrachromosomal genetic elements have been characterized for the archaeal kingdoms Euryarchaeota and Crenarchaeota. Whereas most known euryarchaeal viruses are of the bacteriophage head and tail types, diverse crenarchaeal DNA viruses which exhibit spindle, filamentous, and rod shapes have been characterized (Zillig *et al.,* 1988, 1998). Most of these infect strains of the extremely thermophilic genus *Sulfolobus.* They have been assigned to four novel families: *Fuselloviridae* (SSV1, SSV2, SSV3) (Zillig *et al.,* 1998), *Lipothrixviridae* (SIFV) (Arnold *et al.,* 2000a), *Rudiviridae* (SIRV1, SIRV2) (Prangishvili *et al.,* 1999), and *Guttaviridae* (SNDV) (Arnold *et al.,* 2000b).

The two rudiviruses SIRV1 and SIRV2 have a rod-shaped morphology in which the double-stranded DNA forms a superhelix with a single basic binding protein, with three tail fibers at either end (Prangishvili *et al.,* 1999). The two strands of the linear genomes are covalently linked to form a continuous polynucleotide chain (Prangishvili *et al.,* 1999; Blum *et al.,* 2001). In their original hosts, *Sulfolobus islandicus* strains KVEM10H3 and HVE10/2, respectively, SIRV1 and SIRV2, are present in stable carrier states. However, upon infection of other host strains, SIRV1, but not SIRV2, accrues mutations at a rate of about 10^{-3} substitutions per nucleotide per replication cycle, unprecedented for DNA viruses and approaching mutation rates of the most rapidly varying RNA viruses (Prangishvili *et al.,* 1999). This transient mutation system leads to the selection of conditionally stable SIRV1 variants that after adaptation to the new host replicate with high fidelity.

In order to understand the functions and the functional differences of the two viruses, as well as their evolution-

[1] Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession Nos. AJ344259 and AJ414696.

[2] To whom correspondence and reprint requests should be addressed. Fax: +49.941.943.2403. E-mail: david.prangishvili@biologie.uni-r.de.

0042-6822/01 $35.00

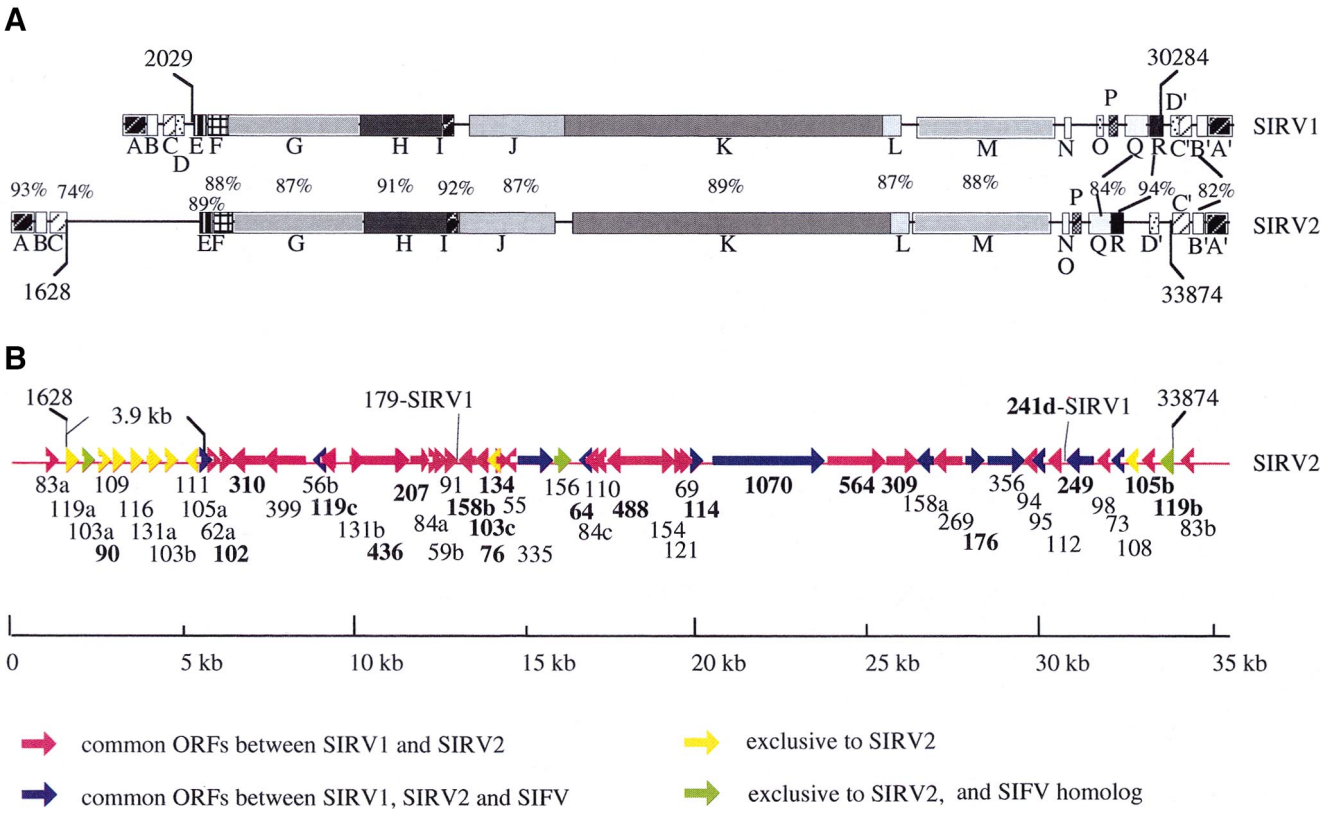

FIG. 1. Comparative nucleotide sequence and predicted ORFs of the genomes of SIRV1 and SIRV2. (A) Regions of similar nucleotide sequence are presented by blocks labeled A to R. The high levels of nucleotide sequence similarity between blocks are indicated. Internal limits of the terminal inverted repeats (nucleotides 2029 and 30,284 in SIRV1 and 1602 and 33,848 in SIRV2) are indicated. (B) The ORF map of SIRV2 is aligned with the nucleotide map in (A). Locations of ORFs which are exclusive to SIRV1 are indicated above the map. Red arrows represent ORFs which have homologs in SIRV1; yellow arrows represent ORFs exclusive to SIRV2; blue arrows represent SIFV viral homologs, one of which, ORF158a, has homologs in different archaeal chromosomes; green arrows represent ORFs exclusive to SIRV2 that are also SIFV homologs. Numbers in boldface type represent ORFs giving matches with ORFs of eukaryal viruses.

ary history including their relationship to other archaeal viruses, the genomes of the conditionally stable variant VIII of SIRV1 and of the stable SIRV2 were sequenced and analyzed.

RESULTS

Genome organization

The linear genomes of SIRV1 and SIRV2 contain 32,312 and 35,502 bp, with inverted terminal repeats (ITRs) of 2029 and 1628 bp, respectively. While the covalently linked termini of SIRV1 were analyzed by chemical sequencing and were shown to be identical and able to form a fully basepaired structure (Blum *et al.,* 2001), about 26 bp at each end of SIRV2 were not determined. Both viruses have the same low G+C content of 25% compared with 33 and 40%, respectively, for the *Sulfolobus* viruses SIFV (Arnold *et al.,* 2000a) and SSV1 (Palm *et al.,* 1991) and 37% for the genome of *Sulfolobus solfataricus* (She *et al.,* 2001).

No sequence heterogeneity was detected except for an ambiguity observed in the genome of the SIRV1 variant VIII. At positions 1075 and 31,233 in the two ITRs, 6 of 33 sequences carried the 12-nucleotide insertion TGAACTGCTAAA.

Sequence comparisons revealed that the viral genomes have similar structures where regions of high sequence conservation are separated by sequences with low or no similarity. In Fig. 1 conserved regions are represented by blocks that are 74 to 94% identical in sequence. The conserved central regions of the genomes contain nine blocks, E to M, varying in length from 293 to 9303 bp and their separation ranges from a 10-bp overlap to a 479-bp gap. The borders between regions of low and high sequence similarity coincide mainly with ORF boundaries.

The regions adjacent to the terminal repeats are composed of smaller conserved blocks interrupted by dissimilar sequences that are rich in repeats. Near the termini of both viral genomes (in blocks A and A′ in Fig. 1) there are three tandem direct repeats, TTTTTTGC, and seven imperfect direct repeats with highly conserved sequences that flank one to three copies of the sequence AAATTCC (Fig. 2). Inverted repeats lie within

```
70          T-AAATTGG AAATTCC AATTTTTTT                    93
118  AAAGTTAAT-AAATTGa  AAATTCC TAAATTCC TAAATTCC AATTATT   163
190  AAAGTTAAT-AAATTGG  AAgTTCC TAAATTCC AATTAATT           227
254  AAA-TTAATAAAATTGa  AAgTTag gAAATTCC TAAATTCC AATAAATT  300
325  AAAGTTAATAAAATTGt  AAATTCC AAAATTCC AATTAATT           364
398          T-AAATTGG  AAATTaC AAAATTCC AAaTATT            427
453  AAAATTgtT-tAATTGt  AAATTCC TAAATTCC AAaTAATT           492

consensus  TxaAATTGg( AAATTCC )n AAtt(A/T) 3-5   n: 1, 2 or 3
```

FIG. 2. Direct repeat sequences at the termini of the genome of SIRV1 and SIRV2. Nucleotide numbers define the start and end positions of each copy in SIRV2. The core sequence is in boldface type and flanking sequences on the right side are italicized. Internal inverted repeats are underlined in the boxed consensus sequence. Capital letters indicate conserved positions, and sequences of direct repeats in the termini of SIRV1 are almost identical to those in SIRV2, except that the sequence of the first 24 nucleotides in the fourth fragment is more conserved in SIRV1 (AAATTAATA-AAATTGG**AAATTCC**T at positions 254–277).

these seven direct repeats (Fig. 2). Our analysis of the partial genome sequence of the lipothrixvirus SIFV (Arnold *et al.,* 2000a) and further sequencing of the ends of the genome reveal that it also carries long terminal inverted repeats. They span at least 800 bp and contain at each end at least seven perfect 27-bp tandem repeat sequences as well as a 70-bp inverted repeat (not shown).

The terminal repeat of SIRV2 spans 1628 bp and contains the conserved blocks A to C. It is 400 bp shorter than that of SIRV1, which is 2029 bp long. This difference probably reflects the finding that insertions have occurred in both terminal sequences during the evolution of SIRV2. Adjacent to the left terminal repeat of SIRV2 lies a 3.9-kb sequence flanked by direct repeats of 57 bp (Fig. 3A). This sequence contains 8 pairs of direct repeats, including one of 103 bp (Fig. 3A, Table 1A). The 103-bp sequence also borders both the left terminal repeat and block E in SIRV1 (Fig. 3A). At the right terminus of SIRV2, the 250-bp-long block D′, which constitutes more than half of the absent 400 bp of the terminal repeat of SIRV1 variant VIII, is separated from the ITR by an insertion of about 500 bp (Fig. 1A). A 475-bp region between blocks L and M of SIRV1 contains four direct repeats including a 64-bp tandem repeat, which constitute 50% of this region (Fig. 3B, Table 1B), and another 21-bp direct repeat flanks the gap between blocks J and K in SIRV2.

Both genomes carry the sequence AAAAAAATTG-GAATTTCCAATTTA 67 nucleotides away from the termini. This sequence resembles a consensus resolution sequence $AAAAAAAN_{7-9}(A/G)TTT(A/T)$ that is present at the corresponding position in poxviral genomes and probably is recognized by poxvirus-encoded Holliday junction resolvases in a course of resolution of replicative intermediates (reviewed in Moss, 1996).

Genome replication

In order to obtain insight into the mechanism of genome replication, replicative intermediates of SIRV1 were studied. DNA isolated from particles of SIRV1 variant VIII, and virus-infected host cells, were digested with *Eco*RI and *Xba*I, which have no recognition sequences inside the ITRs. A Southern blot analysis of the restriction

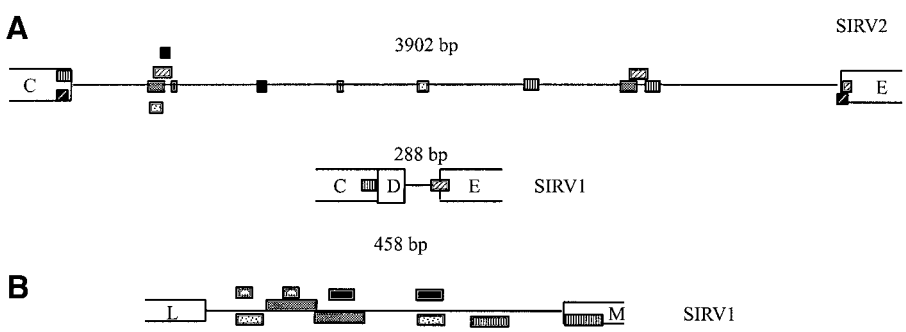

FIG. 3. Schematic presentation of regions rich in direct repeat sequences. (A) In the region between blocks C and E of SIRV2, repeat sequences are indicated by pairs of blocks with identical shading. The repeats range from 33 to 103 bp, and their sequence identities are 75–94%. Sequences similar to these repeats are also indicated between blocks C and E in SIRV1 by boxes with the same shading. The lengths of nucleotide sequences between blocks D and E in SIRV1 and blocks C and E in SIRV2 are indicated. (B) Direct repeats present between blocks L and M in SIRV1 are also represented by pairs of boxes with identical shading.

TABLE 1

Direct Repeat Sequences in the Genomes of SIRV1 and SIRV2

Length (bp)	Mismatches	Positions	Spacing (bp)
A. In the 3.9 kb between blocks C and E in SIRV2			
103[a]	16	1,980; 4,403	2321
85	6	1,942; 4,366	2340
67[b]	11	1,512; 3,858	2380
	9	3,858; 4,476	552
62	6	1,953; 3,321	1307
57	6	1,476; 5,460	3928
39	3	2,013; 2,505	454
33	2	2,060; 2,912	820
B. Between blocks L and M in SIRV1			
64	8	22,760; 22,820	−4
47	0	23,024; 23,144	73
34	3	22,838; 22,952	80
34	7	22,719; 22,952	199
21	0	22,719; 22,779	39

[a] Direct repeats show 75% sequence identity to the SIRV1 genome at nucleotides 1983–2085 (Fig. 2A).

[b] Three copies show 85% sequence identity to the SIRV1 genome at nucleotides 1599–1665 (Fig. 2A).

fragments was performed with the ^{32}P-labeled oligonucleotide 5′-GACGGAAAAGTTTTGGTTCCTCCT-3′, which is complementary to sequences within the ITRs. The results are shown in Fig. 4. For DNA from virus-infected cells, the probe hybridized to the two terminal fragments of the SIRV1 DNA, which are 2049 and 3285 bp long, and to two fragments twice their length. The results indicate the formation of head-to-head and tail-to-tail linked virus DNA molecules in the virus-infected cells. No band was detected with a size of about 5.3 kb, which would have arisen from tail-to-head linked concatemers. We cannot explain the presence of positive signals from weak bands larger than 6.6 kb (Fig. 4).

Analysis of ORFs

SIRV1 and SIRV2 contain about 45 and 54 ORFs, respectively, within the size range 55 to 1070 amino acids, of which 44 are homologous in both genomes. In the map of SIRV2 shown in Fig. 1B red and blue arrows represent these common ORFs. They are more conserved in the central region (identity/similarity 82–99/90–100%) than toward the termini (41–86/69–95%).

Two adjacent ORFs, 241d and 252d of SIRV1, show 41/59% sequence identity/similarity to each other and 41/61% and 65/83% sequence identity/similarity, respectively, to a single homolog, ORF 249, in the SIRV2 genome. It is also likely that ORFs 73, 95, and 105a from both SIRVs, and ORFs 103a and 119b, which are exclusive to SIRV2, have a common origin because they share 24–35/49–55% identity/similarity.

Functions could be assigned to only a few ORFs (Table 2). The N-terminus of ORF134 (AKGHTSRSYSQRYAKPQAKFNAFS) is identical to the N-terminus of the major DNA-binding structural protein isolated from viral particles (Prangishvili *et al.,* 1999). ORF 158b corresponds to 2′-deoxyuridine 5′-triphosphatase (dUTPase) that was characterized for SIRV1 variant II (Prangishvili *et al.,* 1998). ORF 121 shows a high level of sequence similarity to the Holliday junction resolvase of *S. solfataricus* (Kvaratskhelia and White, 2000) and ORFs 356 and 335 carry a sequence motif characteristic for group 1 glycosyl transferases. ORF 55 contains a zinc-finger C2H2 domain.

Significant database matches were also obtained for a further 15 ORFs, common to both genomes, and 3 exclusive to SIRV2, most of which are classified as hypothetical archaeal proteins (Table 2, Fig. 1B). Of these, 12 ORFs shared by both viruses (blue arrows) and 3 ORFs exclusive to SIRV2 (green arrows) are homologous to ORFs of the *Sulfolobus* virus SIFV (Fig. 1B), albeit with a different genomic distribution. Three of these ORFs show significant similarities to their SIFV analogs also at the nucleotide sequence level (Table 3).

Twenty-one ORFs in the SIRV genomes gave positive matches with eukaryal viral genes. Fourteen of these are with ORFs of the *Poxviridae* (Table 4, Fig. 1B), and eight are with the *Amsacta moorei* entomopoxvirus (AmEPV), which also has an A+T-rich genome (82.2%) (Bawden *et*

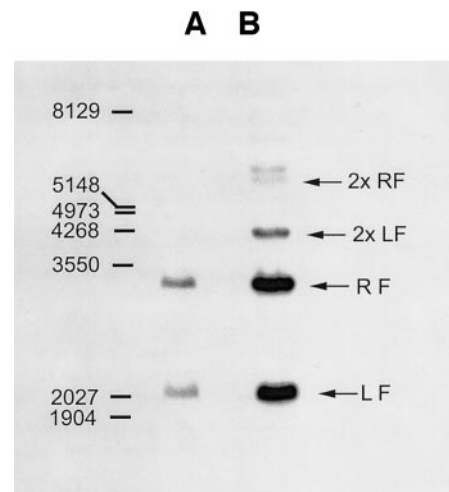

FIG. 4. Dimeric terminal fragments of SIRV1 DNA in SIRV1-infected cells. 0.2 μg DNA isolated from purified SIRV1 variant VIII particles and 15 μg total DNA from cells of *S. islandicus* REN2H1 infected with SIRV1 variant VIII were double-digested with restriction endonucleases *Eco*RI and *Xba*I and processed as described under Materials and Methods. The fragments were detected by hybridization with a ^{32}P-labeled 24-mer oligonucleotide complementary to a portion of the ITR sequence. (A) DNA from virus particles. (B) DNA from virus-infected cells. Lengths of markers in kb are indicated on the left and arrows indicate the left terminal fragment (LF), right terminal fragment (RF), and fragments with twice the length of the terminal fragments.

TABLE 2

Putative Homologs of SIRV ORFs[a]

SIRV2 ORFs	Best matches	aa-aligned (gap residues)	Identity/similarity (%)	*E*-value
207	Hypothetical MJ0377, 170 aa, *Methanococcus jannaschii*	171 (30)	28/47	1×10^{-4}
59b	Hypothetical, 72 aa, *Sulfolobus* pNOB8	36 (0)	50/72	1×10^{-4}
91	Hypothetical c-102, virus SSV1	90 (4)	29/49	8×10^{-7}
158b	dUTPase, 158 aa, virus SIRV1 variant II	158 (0)	90/94	3×10^{-76}
134	Structural protein of SIRVs[b]			
55	Zinc-finger 30C, 777 aa, *Drosophila melanogaster*	30 (0)	40/73	0.009
335	Putative glycosyl transferases group 1, 368 aa, *Pyrococcus abyssi*	160 (27)	27/44	8×10^{-6}
121[c]	Holliday junction resolvase, 143 aa, *S. solfataricus*	79 (6)	37/55	3×10^{-8}
356	Putative glycosyl transferases group 1, 885 aa, *Aquifex aeolicus*	321 (44)	23/42	6×10^{-10}

[a] Homologs which are shared with the SIFV virus are indicated in Fig. 1B including ORFs 335 and 356.
[b] Prangishvili *et al.* (1999).
[c] Homologs of ORF 121 were found in all the sequenced archaeal genomes.

al., 2000). Seven other positive matches were with ORFs of members of the families *Asfarviridae* (African swine fever virus; ASFV) and *Pycodnaviridae* (*Chlorella* viruses). Sequence identities/similarities were in the range 20–57/37–69%, covering stretches of 28 to 437 aa, using a maximum *e*-value of 10 in BLAST searches. Genomic analyses of other archaeal viruses, including the *Sulfolobus* viruses SIFV and SSV1 and the methanogenic phage ψM2, reveal that the SIRVs share many more homologs with poxviruses than do SSV1 and ψM2, while SIFV exhibits a similar number of poxvirus homologs and seven bacteriophage homologs within its larger genome (Table 4).

A careful search for a virus-encoded DNA polymerase which, as for the related eukaryal viruses, should be responsible for DNA replication revealed no conserved motifs characteristic of DNA polymerases. The only database match with the BLAST search was between ORF 399 and the DNA polymerase of the linear mitochondrial plasmid of the fungus *Gelasinospora* (Yuewang *et al.,* 1996) where a 124-amino-acid fragment extending from positions 638 to 761 of the DNA polymerase (987 aa) shows 23/44% identity/similarity.

DISCUSSION

Relationships with other archaeal viruses

Complete sequences so far have been reported for two archaeal viruses, the fusellovirus SSV1, which infects the crenarchaeon *Sulfolobus* (Palm *et al.,* 1991), and the tailed phage ψM2, which infects the euryarchaeon *Methanobacterium thermoautotrophicum* (Pfister *et al.,* 1998). Near complete sequences, excluding terminal regions, have been determined for the linear genomes of the lipothrixviruses TTV1 (EMBL Accession No. X14855; Neumann, 1988) and SIFV (Arnold *et al.,* 2000a). About 45% of the genome of the tailed phage ϕH that infects the euryarchaeon *Halobacterium* has also been sequenced (Stolt, 1993).

As expected, no significant sequence similarities were found between ORFs of the rod-shaped rudiviruses and tailed phages of the euryarchaea, whereas similarities have been found between ORFs of the tailed euryarchaeal phage ψM2 and tailed phages of gram-positive bacteria (Pfister *et al.,* 1998). Moreover, 16 putative ORFs of the rudiviruses have homologs in the *Sulfolobus* lipothrixvirus SIFV, suggesting that the two viral families are

TABLE 3

Comparison of Nucleotide and Amino Acid Sequence Identities for Homologous ORFs in the SIRV2, SIRV1, and SIFV Viral Genomes

	Amino acid identity (%)		Nucleotide identity (%)		G+C content (%)	
SIRV2 ORFs	SIRV2/SIRV1	SIRV2/SIFV	SIRV2/SIRV1	SIRV2/SIFV	SIRV2	SIFV
156	—	85	—	91	27.8	28.8
64	82	73.5	87	75	21.5	34.3
158a	94	55	91	65	25.2	33.8

TABLE 4

A Summary of Positive Database Matches between ORFs of SIRVs and Other Viruses

Archaeal viruses	Genome size (kb)	Number of ORFs	Viral matches[a]	Poxviral matches	Bacteriophage matches
SIRV2	35.5	54	21	14	1
SIFV	40.0	74	34	13	7
SSV1	15.5	31	9	4	2
psi M2	26.1	31	11	3	6

[a] Only eukaryal viral matches and bacteriophage matches are included.

related. Our analysis and further sequencing of terminal regions of the linear SIFV genome (Arnold *et al.,* 2000a) also revealed the presence of large inverted repeats as well as tandem direct repeats in these ITRs, suggesting that the organization of the genomic termini is similar to that of the SIRV viruses.

Relationships with eukaryal viruses

Covalently closed ends and long ITRs are characteristics that SIRV1 and SIRV2 share with eukaryal viruses including poxviruses (Moss, 1996), ASFV (Gonzalez *et al.,* 1986), and *Chlorella* viruses (Zhang *et al.,* 1994). This similarity is reinforced by a detailed analysis of the ITRs of the SIRVs, which revealed imperfect and perfect direct repeats that are also common in the terminal regions of these eukaryal virus genomes (de la Vega *et al.,* 1990; Moss, 1996; Nishida *et al.,* 1999). The number and lengths of these repeats vary in different eukaryal viruses, but similarities occur between the repeat sequences in a given viral family and also between repeats in the same genome, which are assumed to arise from unequal cross-over events (Moss, 1996; Dixon *et al.,* 1990; Nishida *et al.,* 1999). The variable number of core sequences, AAATTCC, in seven imperfect direct repeats in the ITRs of both SIRV genomes (Fig. 2), is also likely to have resulted from unequal cross-over events.

Genome analysis also has revealed other features shared by the SIRVs and eukaryal poxviruses, ASFV, and *Chlorella* viruses. One of these is the location of hotspots for genetic recombination near the genome termini (de la Vega *et al.,* 1990; Nishida *et al.,* 1999). Moreover, conserved genes are concentrated near the center of the archaeal genomes, whereas less conserved genes are located toward the extremities, as occurs for AmEPV poxvirus (Bawden *et al.,* 2000). Finally, BLAST searches revealed that 14 ORFs of the SIRV genomes have predicted homologs in the genomes of poxviruses, ASFV, and *Chlorella* viruses.

The structural similarities between the genomes suggest that the SIRVs and these eukaryal viruses share common mechanisms of DNA replication. Furthermore, the finding of nicks 11 nucleotides from each terminus in about 5% of the SIRV DNA molecules suggests that the replication of the SIRV genomes, like that of the eukaryal poxviruses, is initiated by generating a free 3′-OH near the termini that primes replication (Blum *et al.,* 2001). Here we provide evidence that the subsequent replication process is also similar. The detection of replicative intermediates of viral DNA linked head-to-head and tail-to-tail, similar to those detected in a course of replication of ASFV DNA (Gonsalez *et al.,* 1986; Rojo *et al.,* 1999), is compatible with the self-priming mechanism of replication proposed for poxvirus genomes (Baroudy *et al.,* 1982). The absence of detectable head-to-tail linked concatemers precludes the possibility that SIRV replicates by a rolling circle mechanism via circularization of nicked linear DNA.

The head-to-head and tail-to-tail linked concatemers that form during poxvirus genome replication generate cruciform structures (Baroudy *et al.,* 1982) that are resolved by a virus-encoded Holliday junction resolvase (Stuart *et al.,* 1992; Garcia *et al.,* 2000). A highly conserved sequence, AAAAAAAN$_{7-9}$(A/T)TTT(A/G), present near the termini of the poxviral genomes is recognized by these enzymes in this process (Merchlinsky, 1990). A similar sequence occurs in the terminal regions of both SIRV genomes. Given that the ORFs 121 in the two SIRV genomes encode Holliday junction resolvases (Birkenbihl *et al.,* 2001), the sequence motif is likely to be involved in the resolution of replicative intermediates even though there is little sequence homology between the archaeal and the eukaryal Holliday junction resolvases. The high sequence conservation of the adenosine tracts may reflect their capacity to induce DNA bending (Burkhoff and Tullius, 1987), which could be important for intermediate formation during strand exchange (Merchlinsky, 1990).

In conclusion, the archaeal rudiviruses apparently share characteristics of the linear organization of their genomes and mechanisms of both DNA replication and resolution of replicative intermediates with the eukaryal viruses, indicating a common origin at least of their replication machineries.

Processes involved in evolution of the SIRV viruses

Recombination. Sequence analysis revealed in the two SIRV genomes the existence of well-conserved sequence regions separated by divergent sequences. The major recombination events have occurred within or close to the ITRs. It is likely that a 3.9-kb fragment of the SIRV2 genome, which is flanked by a direct repeat of 57 bp, is a result of insertion into the left ITR of the SIRV1 genome and that about 0.5 kb have been inserted into the right ITR (Fig. 2A). The presence of eight pairs of direct repeats within, and bordering, the former insert, suggests that multiple rearrangements have occurred (Fig. 2A, Table 3A). Evidence for multiple rearrangements were also found outside the ITRs. A region between blocks L and M of SIRV1 contains four direct repeats, which encompass about half of that sequence (Fig. 2B, Table 3B), and a region between blocks J and K in SIRV2 is also flanked by direct repeats.

Gene duplication. DNA duplication followed by divergent evolution of the duplicated genes is an important mechanism for generating new gene functions and has occurred extensively during the evolution of bacterial, archaeal, and eukaryal chromosomes (Labedan and Riley, 1995; Macario *et al.,* 1999; Lodish *et al.,* 1999; Yanai *et al.,* 2000). A few examples are also known for bacterial and eukaryal viruses (Kutter *et al.,* 1996; LaPierre *et al.,*

1999) but the SIRV genomes provide the first evidence for archaeal viruses. For example, the adjacent ORFs 241d and 252d of SIRV1 (Fig. 1) show 41/59% sequence identity/similarity to each other and 41/61% and 65/83% sequence identity/similarity, respectively, to ORF 249 of SIRV2. Thus, ORF 252d of SIRV1 is more closely related to ORF 249 of SIRV2 than to ORF 241d of SIRV1, indicating that gene duplication occurred and then ORF 241d diverged more rapidly than ORF 252d. Evidence of multiple gene duplication was furnished by ORFs 73, 95, and 105a, shared by both SIRVs, and ORFs 103a and 119b, exclusive to SIRV2. All show 24–35/49–55% identity/similarity to one another, indicative of a common origin. They also show sequence similarity to ORF 112 (No. 14) of SIFV.

Horizontal gene transfer. Evidence for horizontal gene transfer arose from a comparative study of the 16 putative ORFs common to the SIRV and SIFV genomes. Three common ORFs, 64, 156, and 158a, show a higher level of similarity of nucleotide sequence than amino acid sequence, a property which generally indicates that the sequence constitutes a *cis*-element. Of these, ORF 156 of SIRV2, which has no homolog in SIRV1, and ORF 156 (No. 44) of SIFV are exceptional in that they show a nucleotide sequence identity as high as 91%, compared with an amino acid sequence identity of only 85% (Table 3). Furthermore, the G+C content of the ORFs is close despite the higher average G+C content of the SIFV genome (Table 3). Given the major differences between the SIRVs and SIFV in DNA sequence, genome organization, and morphology, we infer that the gene in SIRV2 has been horizontally transferred from an SIFV-like virus.

Of other ORFs with high nucleotide sequence identities, ORF 64 from the SIRVs and ORF 64 (No. 62) from SIFV (Table 3) are likely to constitute a highly conserved *cis*-element, while ORFs 158a from the SIRVs and ORF 158 (No. 65) from SIFV probably produce a conserved protein that is also encoded in archaeal genomes (Fig. 1B).

Displacement of viral enzymes by host enzymes. Evidence for a common evolutionary origin of the SIRV viruses and the eukaryal poxviruses, ASFV and *Chlorella* viruses, correlates with the functional similarities of these viruses and the homology of ORFs as shown in Table 4. However, there is also evidence for the presence in the SIRVs of genes with higher sequence similarity to cellular genes than to homologs in related viruses. For example, the dUTPases of the SIRVs, the poxviruses, and the *Chlorella* viruses are more closely related to the corresponding enzymes of their hosts than to one another (Prangishvili *et al.,* 1998; Baldo and McClure, 1999). This suggests that the viral genes have been replaced by genes from the host chromosome during coevolution of the viruses and their respective hosts. In contrast, phylogenetic analyses of the dUTPase of ASFV virus indicate that it clusters with dUTPases from different herpes viruses and is only distantly related to eukaryal cellular dUTPases (Baldo and McClure, 1999).

The Holliday junction resolvases of SIRVs also constitute a probable example of the exchange of viral genes by their cellular homologs. The viral genes have close relatives in the chromosome of *S. solfataricus* and other archaea but show no significant sequence similarity to the corresponding genes from poxviruses. In contrast, some poxviral genes have homologs in bacteria and fungal mitochondria and may derive from exchange with mitochondrial genomes (Garcia *et al.,* 2000; Lilley and White, 2000).

CONCLUDING REMARKS

One of the main aims of comparative sequence analysis of the genomes of viruses SIRV1 and SIRV2 was to shed light on the mechanism which enables SIRV1, but not the stable SIRV2, to adapt to new hosts by extensive accumulation of point mutations leading to the production of viable variants. The results of the sequence comparisons suggest that the ability to induce variation is either a positive feature or a deficiency of SIRV1. Most of the unique ORFs of SIRV2, e.g., that putatively transferred from an SIFV-like virus (ORF156), and especially the eight ORFs clustered in the 3.9-kb insert, could be later acquisitions and may contribute to the control of replication fidelity. Conversely, the two SIRV1-specific ORFs, 179 and 241d, could induce variability of this virus. Selective deletion or insertion of these unique ORFs from, or into, the viral genomes should provide insight into the regulation of the fidelity of replication and explain the puzzling differences in viral adaptation to alternative hosts.

Comparative genome analysis also revealed potential sites where insertions could occur in the viral genomes without affecting basic viral functions. This should provide a basis for the development of transformation vectors based on DNAs of SIRV1 and SIRV2, thereby enriching the small number of genetic tools currently available for extremely thermophilic archaea.

Viruses evolve rapidly and the resulting genetic diversity makes it difficult to trace evolutionary lineages by sequence comparison. Previously, we have suggested a phylogenetic relationship between the archaeal rudivirus SIRV1 and the eukaryal poxviruses, ASFV and the *Chlorella* viruses (Blum *et al.,* 2001). This hypothesis was based mainly on similarities in the structure of the terminal regions of the genome of SIRV1 and the eukaryal viruses. The analysis of the genome sequences of the rudiviruses SIRV1 and SIRV2 and of a mode of replication of their DNA described here provide strong support for this contention.

It has been argued that common ancestry of viruses from different domains of life could not be explained by spreading of viruses between the domains (Zillig *et al.,* 1998; Prangishvili *et al.,* 2001). This would require multi-

ple adaptive changes in order to surmount biochemical barriers, including the incompatibility of transcription systems, as well as complex adaptations to very diverse environments. A more probable scenario is that common ancestors of such viruses existed prior to the divergence of the three domains (Zillig *et al.,* 1996, 1998; Hendrix, 1999). If this is true, then studying viral evolution should yield important insights into the early stages of evolution.

MATERIALS AND METHODS

Sequencing of the SIRV1 genome

DNA isolated from purified particles of SIRV1 variant VIII (Zillig *et al.,* 1994) was fragmented by sonification. After formation of blunt ends with a mixture of T4 DNA polymerase and Klenow enzyme, fragments of 400 to 1000 bp, or 1000 to 2000 bp, were isolated by preparative gel electrophoresis and ligated into 11 different multiplex vectors (Church and Kieffer-Higgins, 1988). Up to six different clones were combined and 500 ng of these mixtures was sequenced enzymatically (Thermosequenase Kit, Amersham) essentially as described by Chee (1991). Sequence patterns were detected by repeated hybridization with 3′-digoxigenated oligonucleotides and visualized by phosphate-triggered chemiluminescence of CDP-Star. The hybridizations and antibody-based hybrid detection reactions were performed in a "membrane processor." This automated device was able to process up to 80 sequencing reactions in parallel with an average reading length of about 220 nucleotides (developed by JOBO, Gummersbach, Germany). The data from 1450 sequencing reactions were assembled using the program Seqman (DNAStar). The produced a 10- to 11-fold overall coverage of the SIRV1 genome. Gaps were closed by primer walking. The terminal regions were sequenced by chemical degradation, as described previously (Blum *et al.,* 2001).

Sequencing of the SIRV2 genome

DNA isolated from purified particles of SIRV2 was sequenced using a shotgun cloning method as described previously (She *et al.,* 1998). DNA sequencing was done in a MegaBACE 1000 Sequenator (Amersham–Pharmacia) using dye-terminator chemistry. Primers for gap filling and sequence polishing were designed by the "Primers for Mac" program (Resnick, Richard, Primer V1.0. Ashland, MA) and sequence data were assembled by Sequencher 3.0 (Gene Code, Ann Arbor, MI).

Sequence analysis

ORFs were identified using Sequencher 3.0 and were searched for matches in NCBI databases using BLAST (Altschul *et al.,* 1997). Nucleotide and amino acid sequences were compared with BLAST 2 (Tatiana and Thomas, 1999). Repeat sequences were located using the program Luna (Brügger *et al.,* unpublished data).

DNA analysis

DNA from cells of *S. islandicus* REN2H1 24 h postinfection with SIRV1 variant VIII at a m.o.i. of 10 was purified (Zillig *et al.,* 1994). DNAs from purified SIRV1 particles and from virus-infected host cells were digested with restriction endonucleases *Eco*RI and *Xba*I. DNA fragments were analyzed on 1% agarose gels. An oligonucleotide probe was labeled with [^{32}P]ATP by T4 polynucleotide kinase and Southern hybridization was performed using standard procedures (Sambrook *et al.,* 1989).

ACKNOWLEDGMENTS

Parts of this work were funded by EU in the frame of its Biotechnology Programme, Project "Extremophiles as Cell Factories," and by grants from the Novo Nordisk Foundation, the Danish Research Council for Natural Science, the Bundesministerium für Bildung und Forschung, and the Deutsche Forschungsgemeinschaft (Pr 663/1-1). Thanks are extended to A. Kessler and M. Häring for assistance in studies on replicative intermediates. D.P. gratefully acknowledges the support of Karl O. Stetter.

REFERENCES

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402.

Arnold, H. P., Zillig, W., Ziese, U., Holz, I., Crosby, M., Utterback, T., Weidmann, J. F., Kristjanson, J. K., Klenk, H. P., Nelson, K. E., and Fraser, C. M. (2000a). A novel lipothrixvirus, SIFV, of the extremely thermophilic crenarchaeon *Sulfolobus*. *Virology* **267,** 252–266.

Arnold, H. P., Ziese, U., and Zillig, W. (2000b). SNDV, a novel virus of the extremely thermophilic and acidophilic archaeon *Sulfolobus*. *Virology* **272,** 409–416.

Baldo, A. M., and McClure, M. A. (1999). Evolution and horizontal transfer of dUTPase-encoding genes in viruses and their hosts. *J. Virol.* **73,** 7710–7721.

Baroudy, D., Venkatesan, S., and Moss, B. (1982). Incompletely base-paired flip-flop terminal loops link the two DNA strands of the vaccinia virus genomes into one uninterrupted polynucleotide chain. *Cell* **28,** 315–324.

Bawden, A. L., Glassberg, K. J., Diggans, J., Shaw, R., Farmerie, W., and Moyer, R. W. (2000). Complete genomic sequence of the *Amsacta moorei* entomopox virus: Analysis and comparison with other pox viruses. *Virology* **274,** 120–139.

Birkenbihl, R. P., Neef, K., Prangishvili, D., and Kemper, B. (2001). Holliday junction resolving enzymes of archaeal viruses SIRV1 and SIRV2. *J. Mol. Biol.* **309,** 1067–1076.

Blum, H., Zillig, W., Mallok, S., Domdey, H., and Prangishvili, D. (2001). The linear genome of the archaeal virus SIRV1 has features in common with genomes of eukaryal viruses. *Virology* **281,** 6–9.

Burkhoff, A. M., and Tullius, T. D. (1987). The unusual conformation adopted by the adenine tracts in kinetoplast DNA. *Cell* **48,** 935–943.

Chee, F. (1991). Enzymatic multiplex DNA sequencing. *Nucleic Acids Res.* **19,** 3301–3305.

Church, G. M., and Kieffer-Higgins, S. (1988). Multiplex DNA sequencing. *Science* **240,** 185–188.

de la Vega, I., Vinuela, E., and Blasco, R. (1990). Genetic variation and

multigene families in African swine fever virus. *Virology* **179,** 234–246.

Dixon, L. K., Bristow, C., Wilkinson, P. J., and Sumption, K. J. (1990). Identification of a variable region of the African swine fever virus genome that has undergone separate DNA rearrangements leading to expansion of minisatellite-like sequences. *J. Mol. Biol.* **216,** 677–688.

Garcia, A. D., Aravind, L., Koonin, E. V., and Moss, B. (2000). Bacterial-type DNA Holliday junction resolvases in eukaryotic viruses. *Proc. Natl. Acad. Sci. USA* **97,** 8926–8931.

Gonzalez, A., Talavera, A., Almendral, J. M., and Vinuela, E. (1986). Hairpin loop structure of African swine fever virus DNA. *Nucleic Acids Res.* **14,** 6835–6844.

Gropp, F., Palm, P., and Zillig, W. (1992). The immunity-conferring plasmid pHL from *Halobacterium salinarium* phage ϕH: Nucleotide sequence and transcription. *Virology* **190,** 45–54.

Hendrix, R. W. (1999). The long evolutionary reach of viruses. *Curr. Biol.* **9,** R914–R917.

Kutter, E., Gachechiladze, K., Poglazov, A., Marusich, E., Shneider, M., Aronsson, P., Napuli, A., Porter, D., and Mesyanzhinov, V. (1996). Evolution of T4-related phages. *Virus Genes* **11,** 285–297.

Kvaratskhelia, M., and White, M. F. (2000). Two Holliday junction resolving enzymes in *Sulfolobus solfataricus*. *J. Mol. Biol.* **297,** 923–932.

Labedan, B., and Riley, M. (1995). Widespread protein sequence similarities: Origins of *Escherichia coli* genes. *J. Bacteriol.* **177,** 1585–1588.

LaPierre, L. A., Holzschu, D. L., Bowser, P. R., and Casey, J. W. (1999). Sequence and transcriptional analyses of the fish retroviruses walleye epidermal hyperplasia virus types 1 and 2: Evidence for a gene duplication. *J. Virol.* **73,** 9393–9403.

Lilley, D. M. J., and White, M. (2000). Resolving the relationships of resolving enzymes. *Proc. Natl. Acad. Sci. USA* **97,** 9351–9353.

Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., and Darnell, J. (Eds.) (1999). "Molecular Cell Biology," 4th ed., pp. 229–230. Scientific American.

Macario, A. J., Lange, M., Ahring, B. K., and Convey De Macario, E. C. (1999). Stress genes and proteins in the archaea. *Microbiol. Mol. Biol. Rev.* **63,** 923–967.

Merchlinsky, M. (1990). Mutational analysis of the resolution sequence of vaccinia virus DNA: Essential sequence consists of two separate AT-rich regions highly conserved among pox viruses. *J. Virol.* **64,** 5029–5035.

Moss, B. (1996). Poxviruses and their replication. *In* "Field's Virology" (B. N. Fields, D. M. Knipe, and P. M. Howley, Eds.), pp. 2637–2671. Lippincott–Raven, New York.

Neumann, H. (1988). "Struktur, Funktion und Variabilität eines archaebacteriellen Genomes," Ph.D. dissertation, Ludwig-Maximilians-Universität München, Munich, Germany.

Nishida, K., Kimura, Y., Kawasaki, T., Fujie, M., and Yamada, T. (1999). Genetic variation of *Chlorella* viruses: Variable regions localized on the CVK2 genomic DNA. *Virology* **255,** 376–384.

Palm, P., Schleper, C., Grampp, B., Yeats, S., McWilliam, P., Reiter, W. D., and Zillig, W. (1991). Complete nucleotide sequence of the virus SSV1 of the archaebacterium *Sulfolobus shibatae*. *Virology* **185,** 242–250.

Pfeiffer, P., Palm, P., and Schleifer, K.-H. (Eds.) (1994). "Molecular Biology of Archaea." Gustav Fischer Verlag, Stuttgart.

Pfister, P., Wasserfallen, A., Stettler, R., and Leisinger, T. (1998). Molecular analysis of *Methanobacterium* phage ψM2. *Mol. Microbiol.* **30,** 233–244.

Prangishvili, D., Klenk, H. P., Jakobs, G., Schmiechen, A., Hanselmann, C., Holz, I., and Zillig, W. (1998). Biochemical and phylogenetic characterization of the dUTPase from the archaeal virus SIRV. *J. Biol. Chem.* **273,** 6024–6029.

Prangishvili, D., Arnold, H. P., Gotz, D., Ziese, U., Holz, I., Kristjansson, J. K., and Zillig, W. (1999). A novel virus family, the *Rudiviridae:* Structure, virus–host interactions and genome variability of the *Sulfolobus* viruses SIRV1 and SIRV2. *Genetics* **152,** 1387–1396.

Prangishvili, D., Stedman, K., and Zillig, W. (2001). Viruses of the extremely thermophilic archaeon *Sulfolobus*. *Trends Microbiol.* **9,** 39–43.

Rojo, G., Garcia-Beato, R., Vinuela, E., Salas, M. L., and Salas, J. (1999). Replication of African swine fever virus in infected cells. *Virology* **257,** 524–536.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). "Molecular Cloning: A Laboratory Manual," 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

She, Q., Phan, H., Garrett, R. A., Albers, S.-V., Stedman, K. M., and Zillig, W. (1998). Genetic profile of pNOB8 from *Sulfolobus:* The first conjugative plasmid from an archaeon. *Extremophiles* **2,** 417–425.

She, Q., Singh, R. K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M. J., Chan-Weiher, C. C.-Y., Clausen, I. G., *et al.* The complete genome of the Crenarchaeote *Sulfolobus solfataricus* P2. *Proc. Natl. Acad. Sci. USA* **98,** 7835–7840.

Stolt, P. (1993). "Studies on the Genetics and Regulation of Transcription in the *Halobacterium salinarium* Phage ϕH," Dissertation der Fakultät für Chemie, Ludwig-Maximilians-Universität, München, Munich, Germany.

Stuart, D., Ellison, K., Graham, K., and McFadden, G. (1992). *In vitro* resolution of pox virus replicative intermediates into linear minichromosomes with hairpin termini by a virally induced Holliday junction endonuclease. *J. Virol.* **66,** 1551–1563.

Tatiana, A. T., and Thomas, L. M. (1999). Blast 2 sequences — A new tool for comparing protein and nucleotide sequences. *FEMS Microbiol. Lett.* **174,** 247–250.

Upton, C., and McFadden, G. (1986). DNA sequence homology between the terminal inverted repeats of Shope fibroma virus and an endogenous cellular plasmid species. *Mol. Cell. Biol.* **6,** 265–276.

Upton, C., Macen, J. L., Maranchuk, R. A., Delange, A. M., and McFadden, G. (1988). Tumorigenic pox viruses: Fine analysis of the recombination junctions in malignant rabbit fibroma virus, a recombinant between Shope fibroma virus and myxoma virus. *Virology* **166,** 229–239.

Woese, C. R., and Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Natl. Acad. Sci. USA* **74,** 5088–5090.

Yanai, I., Camacho, C. J., and DeLisi, C. (2000). Predictions of gene family distributions in microbial genomes: Evolution by gene duplication and modification. *Phys. Rev. Lett.* **85,** 2641–2644.

Yuewang, W., Yang, X., and Griffiths, A. J. F. (1996). Structure of a *Gelasinopora* linear plasmid closely related to the kalilo plasmid of *Neurospora intermedia*. *Curr. Genet.* **29,** 150–158.

Zhang, Y., Strasser, P., Grabherr, R., and Van Etten, J. L. (1994). Hairpin loop structure at the termini of the *Chlorella* virus PBCV-1 genome. *Virology* **202,** 1079–1082.

Zillig, W., Arnold, H. P., Holz, I., Prangishvili, D., Schweier, A., Stedman, K., She, Q., Phan, H., Garrett, R. A., and Kristjansson, J. K. (1998). Genetic elements in the extremely thermophilic archaeon *Sulfolobus*. *Extremophiles* **2,** 131–140.

Zillig, W., Kletzin, A., Schleper, C., Holz, I., Janekovic, D., Hain, J., Lanzendorfer, M., and Kristjansson, J. (1994). Screening for *Sulfolobales,* their plasmids, and their viruses in Icelandic sulfataras. *Syst. Appl. Microbiol.* **16,** 609–628.

Zillig, W., Prangishvili, D., Schleper, C., Elferink, M., Holz, I., Albers, S., Janekovic, D., and Götz, D. (1996). Viruses, plasmids and other genetic elements of thermophilic and hyperthermophilic archaea. *FEMS Microbiol. Rev.* **18,** 225–236.

Zillig, W., Reiter, W. D., Palm, P., Gropp, F., Neumann, H., and Rettenberger, M. (1988). Viruses of archaebacteria. *In* "The Bacteriophages" (R. Calendar, Ed.), Vol. 1, pp. 517–568. Plenum, New York/London.